DISK CONTROL SYSTEM, DISK CONTROL APPARATUS, DISK SYSTEM AND CONTROL METHOD THEREOF

Publication number: JP2003323261
Publication date: 2003-11-14
Inventor: KANAI HIROKI; KANEKO SEIJI
Applicant: HITACHI LTD
Classification:
- international: ***G06F12/08; G06F3/00; G06F3/06; G06F13/00; G06F13/12;*** **G06F12/08; G06F3/00; G06F3/06; G06F13/00; G06F13/12;** (IPC1-7): G06F3/06; G06F12/08; G06F13/12
- european:
Application number: JP20020126885 20020426
Priority number(s): JP20020126885 20020426

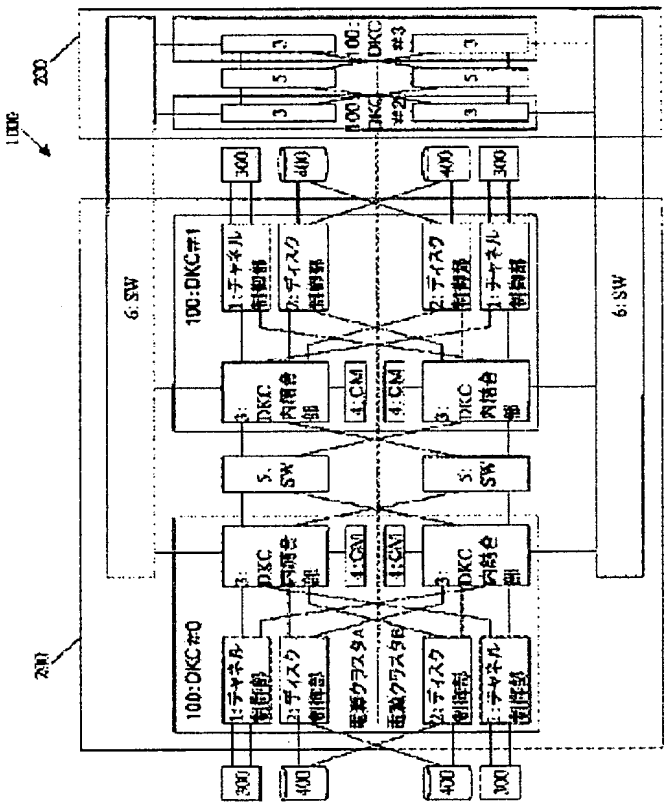
Also published as:
US6961788 (B2)
US2003204649 (A1)

Report a data error here

Abstract of **JP2003323261**

PROBLEM TO BE SOLVED: To provide a control apparatus for effectively configuring with an identical architecture from a small scale configuration to super large scale configuration. SOLUTION: A disk control unit comprises one or a plurality of channel control units having an interface with a host computer, one or a plurality of disk control units having an interface with a disk apparatus, and an internal connection unit for connecting a cache memory unit for temporarily storing data to be read/written from/to the disk apparatus, a channel control unit and a disk control unit. A first connection portion for connecting internal connection portions of disk control units for reading/writing data inside each of the disk control apparatuses, and a second connection portion for connecting internal connection portions of the disk control units for transferring data straddling the plurality of disk control apparatuses.

COPYRIGHT: (C)2004,JPO



Data supplied from the ***esp@cenet*** database - Worldwide

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2003－323261
（P2003－323261A）
(43)公開日　平成15年11月14日(2003.11.14)

(51)Int.Cl.'		識別記号	FI		テーマコード'(参考)
G06F	3/06	301	G06F	3/06	301B　5B005
		302			302A　5B014
					302B　5B065
	12/08	501		12/08	501E
		557			557

審査請求　未請求　請求項の数16　ＯＬ　（全 10 頁）　最終頁に続く

(21)出願番号　特願2002－126885(P2002－126885)

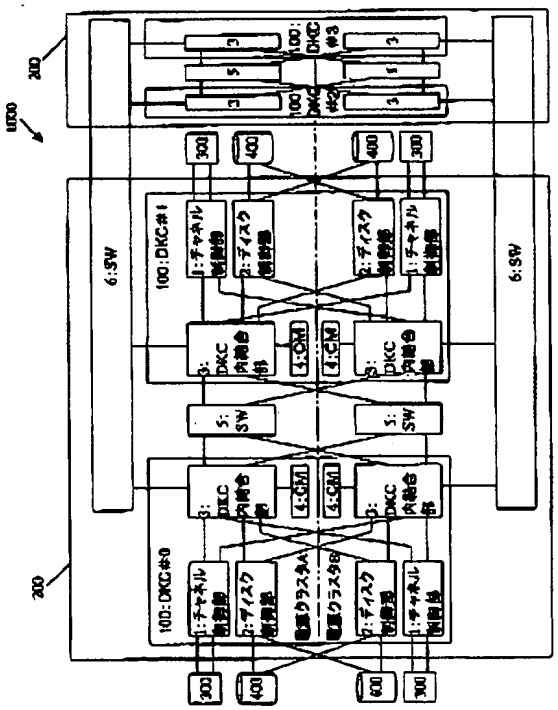
(22)出願日　平成14年４月26日(2002.4.26)

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地
(72)発明者　金井　宏樹
神奈川県小田原市中里322番地２号　株式会社日立製作所ＲＡＩＤシステム事業部内
(72)発明者　金子　誠司
神奈川県小田原市中里322番地２号　株式会社日立製作所ＲＡＩＤシステム事業部内
(74)代理人　100071283
弁理士　一色　健輔　（外４名）

最終頁に続く

(54)【発明の名称】　ディスク制御システム、ディスク制御装置、ディスクシステム、及びその制御方法

(57)【要約】　（修正有）
【課題】小規模な構成から超大規模な構成まで同一のアーキテクチャで効率よく構成することを可能とする制御装置を提供する。
【解決手段】ディスク制御ユニットは、ホストコンピュータとのインターフェースを有する一または複数のチャネル制御部と、ディスク装置とのインターフェースを有する一または複数のディスク制御部と、ディスク装置にリード／ライトされるデータを一時的に格納するキャッシュメモリ部とチャネル制御部とディスク制御部とを相互に接続する内部結合部とを備え、各ディスク制御装置の内部において、データをリード／ライトすべく、各ディスク制御ユニットの内部結合部を相互に結合する第一の結合部と、複数のディスク制御装置に跨り、データを転送すべく、各ディスク制御ユニットの内部結合部を相互に結合する第二の結合部とを備えたものとする。



1

【特許請求の範囲】

【請求項１】 複数のディスク制御ユニットを有するディスク制御装置を複数備えたディスク制御システムにおいて、

前記ディスク制御ユニットは、

ホストコンピュータとのインターフェースを有する一または複数のチャネル制御部と、

ディスク装置とのインターフェースを有する一または複数のディスク制御部と、

前記ディスク装置にリード／ライトされるデータを一時的に格納するキャッシュメモリ部と前記チャネル制御部と前記ディスク制御部とを相互に接続する内部結合部と、

を備えており、

前記各ディスク制御装置の内部において、データをリード／ライトすべく、前記各ディスク制御ユニットの前記内部結合部を相互に結合する第一の結合部と、

複数の前記ディスク制御装置に跨り、データを転送すべく、前記各ディスク制御ユニットの前記内部結合部を相互に結合する第二の結合部と、

を備えたことを特徴とするディスク制御システム。

【請求項２】 前記ディスク制御装置は二つの前記ディスク制御ユニットを有しており、前記第一の結合部は、該二つのディスク制御ユニットの前記内部結合部を相互に結合することを特徴とする請求項１に記載のディスク制御システム。

【請求項３】 前記第一の結合部又は前記第二の結合部は、メモリバス用スイッチで構成されることを特徴とする請求項１に記載のディスク制御システム。

【請求項４】 前記第一の結合部は、データ伝送用のケーブルで構成されることを特徴とする請求項１に記載のディスク制御システム。

【請求項５】 前記各ディスク制御装置の内部において、共通の電源から給電される前記各ディスク制御ユニットを前記第一の結合部は結合することを特徴とする請求項１に記載のディスク制御システム。

【請求項６】 複数の前記ディスク制御装置に跨り、データを転送すべく、前記各ディスク制御ユニットの前記第一の結合部を相互に接続することを特徴とする請求項１に記載のディスク制御システム。

【請求項７】 請求項１に記載のディスク制御システムと、前記ディスク制御システムの前記チャネル制御部に対して接続される、データの授受用のホストコンピュータとを備えることを特徴とするディスクシステム。

【請求項８】 請求項１に記載のディスク制御システムと、前記ディスク制御システムの前記ディスク制御部に対して接続される、データ格納用のディスク装置とを備えることを特徴とするディスクシステム。

【請求項９】 複数のディスク制御ユニットを有するディスク制御装置において、

2

前記ディスク制御ユニットは、

ホストコンピュータとのインターフェースを有する一または複数のチャネル制御部と、

ディスク装置とのインターフェースを有する一または複数のディスク制御部と、

前記ディスク装置にリード／ライトされるデータを一時的に格納するキャッシュメモリ部と前記チャネル制御部と前記ディスク制御部とを相互に接続する内部結合部と、

を備えており、

前記各ディスク制御装置の内部において、データをリード／ライトすべく、前記各ディスク制御ユニットの前記内部結合部を相互に結合する第一の結合部を備えて、

前記各ディスク制御ユニットの前記内部結合部は、複数の前記ディスク制御装置に跨り、データを転送すべく、第二の結合部によって、相互に結合されることを特徴とするディスク制御装置。

【請求項１０】 前記ディスク制御装置は二つの前記ディスク制御ユニットを有しており、前記第一の結合部は、該二つのディスク制御ユニットの前記内部結合部を相互に結合することを特徴とする請求項９に記載のディスク制御装置。

【請求項１１】 前記第一の結合部又は前記第二の結合部は、メモリバス用スイッチで構成されることを特徴とする請求項９に記載のディスク制御装置。

【請求項１２】 前記第一の結合部は、データ伝送用のケーブルで構成されることを特徴とする請求項９に記載のディスク制御装置。

【請求項１３】 共通の電源から給電される前記各ディスク制御ユニットを前記第一の結合部は結合することを特徴とする請求項９に記載のディスク制御装置。

【請求項１４】 複数の前記ディスク制御装置に跨り、データを転送すべく、前記各ディスク制御ユニットの前記第一の結合部が相互に接続されることを特徴とする請求項９に記載のディスク制御装置。

【請求項１５】 ホストコンピュータとのインターフェースを有する一または複数のチャネル制御部と、ディスク装置とのインターフェースを有する一または複数のディスク制御部と、前記ディスク装置にリード／ライトされるデータを一時的に格納するキャッシュメモリ部と前記チャネル制御部と前記ディスク制御部とを相互に接続する内部結合部とを備えた前記ディスク制御ユニットを複数有するディスク制御装置におけるデータ通信の制御方法であって、

前記各ディスク制御装置の内部において、前記各ディスク制御ユニットの前記内部結合部を相互に結合する第一の結合部により、データをリード／ライトするとともに、

複数の前記ディスク制御装置に跨り、前記各ディスク制御ユニットの前記内部結合部を相互に結合する第二の結

合部によって、データを転送することを特徴とするディスク制御装置におけるデータ通信の制御方法。

【請求項１６】 前記ディスク制御装置は二つの前記ディスク制御ユニットを有しており、前記第一の結合部は、該二つのディスク制御ユニットの前記内部結合部を相互に結合することを特徴とする請求項１５に記載のディスク制御装置におけるデータ通信の制御方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】この発明は、ディスク制御システム、ディスク制御装置、ディスクシステム、及びその制御方法に関する。

【０００２】

【従来の技術】半導体記憶装置を記憶媒体とするコンピュータの主記憶のＩ／Ｏ性能に比べて、磁気ディスクを記憶媒体とするディスクサブシステムのＩ／Ｏ性能は３～４桁程度低く、従来からこの差を縮めること、すなわちサブシステムのＩ／Ｏ性能を向上させる努力がなされている。サブシステムのＩ／Ｏ性能を向上させるための１つの方法として、複数の磁気ディスク装置でサブシステムを構成し、データを複数の磁気ディスク装置に格納する、いわゆるディスクシステムと呼ばれるシステムが知られている。このような技術を開示したものとして、特開２００１－２５６００３号公報がある。同公報中の図４に示す技術では、スイッチを用いた相互結合網を介して間接的に、ホストコンピュータ５０が全てのディスク制御装置４に接続されている。

【０００３】しかしながら、複数のディスク制御装置を１つのディスク制御装置として運用するためには、相互結合網を構成するスイッチ内に、そのスイッチに接続された全てのディスク制御装置のデータが、どのディスク制御装置に格納されているかを示すマップを持つ必要があり、ホストコンピュータからアクセス要求があった場合、スイッチにおいてコマンドを解析し、要求データを格納しているディスク制御装置に割り振る機能が必要となる。この場合、従来のチャネルＩＦ部でのコマンド解析に加え、その上に繋がるスイッチにおいてもコマンドを解析する必要があるため、ホストコンピュータがディスク制御装置に直接接続されている場合に比べ、性能が低下するという問題がある。

【０００４】そこで、この特開２００１－２５６００３号公報に開示された発明では、同公報の図１や図８に示されるように、相互結合網を介して、全てのチャネルＩＦ部及びディスクＩＦ部から、全ての共有メモリ部あるいは全てのキャッシュメモリ部へアクセス可能な構成となっている。

【０００５】このような技術により、小規模な構成から超大規模な構成まで、同一の高機能・高信頼性のアーキテクチャで対応可能であって、スケーラビリティのある構成のディスク制御装置を提供できる。

【０００６】

【発明が解決しようとする課題】しかしながら、前述した従来の技術にあっては、データの転送やリード／ライトの処理の効率が未だ不十分である。場合によっては、アクセスに関し、論理的な競合により、相互結合網の効率が５０％以下にまで落ち込んでしまうのである。これを解決しようとすると、広帯域化を図る必要があるが、高価格化を招く。

【０００７】本発明は、このような課題に鑑みてなされたもので、ディスク制御システム、ディスク制御装置、ディスクシステム、及びその制御方法を提供することを目的とする。

【０００８】

【課題を解決するための手段】前記目的を達成すべく、本発明の主たる発明のディスク制御システムでは、複数のディスク制御ユニットを有するディスク制御装置を複数備えたディスク制御システムにおいて、前記ディスク制御ユニットは、ホストコンピュータとのインターフェースを有する一または複数のチャネル制御部と、ディスク装置とのインターフェースを有する一または複数のディスク制御部と、前記ディスク装置にリード／ライトされるデータを一時的に格納するキャッシュメモリ部と前記チャネル制御部と前記ディスク制御部とを相互に接続する内部結合部とを備えており、前記各ディスク制御装置の内部において、データをリード／ライトすべく、前記各ディスク制御ユニットの前記内部結合部を相互に結合する第一の結合部と、複数の前記ディスク制御装置に跨り、データを転送すべく、前記各ディスク制御ユニットの前記内部結合部を相互に結合する第二の結合部とを備えたこととする。

【０００９】その他、本願が開示する課題、及びその解決方法は、発明の実施形態の欄及び図面により明らかにされる。

【００１０】

【発明の実施の形態】本明細書の記載により、少なくとも次のことが明らかにされる。前記ディスク制御装置は二つの前記ディスク制御ユニットを有しており、前記第一の結合部は、該二つのディスク制御ユニットの前記内部結合部を相互に結合することとしてもよい。

【００１１】前記第一の結合部又は前記第二の結合部は、メモリバス用スイッチで構成されることとしてもよい。

【００１２】また、前記第一の結合部は、データ伝送用のケーブルで構成されることとしてもよい。

【００１３】さらに、前記各ディスク制御装置の内部において、共通の電源から給電される前記各ディスク制御ユニットを前記第一の結合部は結合することとしてもよい。

【００１４】さらにまた、複数の前記ディスク制御装置に跨り、データを転送すべく、前記各ディスク制御ユニ

ットの前記第一の結合部を相互に接続することとしてもよい。

【００１５】

【実施例】本発明に係る実施例につき、図面を参照して説明する。ディスク制御システム１０００は、複数のディスク制御装置２００を備えている。各ディスク制御装置２００は、複数、好ましくは二つのディスク制御ユニット（ＤＫＣとも称する）１００（ＤＫＣ＃０乃至ＤＫＣ＃３）を有する。

【００１６】これら各ディスク制御ユニット１００は、チャネル制御部１と、ディスク制御部２と、ＤＫＣ内結合部（内部結合部）３とを備える。例えば、ＤＫＣ内結合部は相互結合網で構成される。チャネル制御部１は、ホストコンピュータ３００とのインターフェースを有する。ディスク制御部２は、ディスク装置４００とのインターフェースを有する。ＤＫＣ内結合部３は、ディスク装置４００にリード／ライトされるデータを一時的に格納するキャッシュメモリ（ＣＭ）４とチャネル制御部１とディスク制御部２とを相互に接続する。

【００１７】さらに、各ディスク制御ユニット１００は、障害対策で電源系統別に二重化されている。ペアの電源クラスタＡ，Ｂそれぞれが、チャネル制御部１と、ディスク制御部２と、ＤＫＣ内結合部３とを備える。ＤＫＣ内結合部３は、他方の電源クラスタのチャネル制御部１及びディスク制御部２、並びに、他方の電源クラスタ側のＳＷ５と接続している。また、ディスク制御部２は、他方の電源クラスタ側のディスク装置４００とも接続されている。

【００１８】各ディスク制御装置２００は、各ディスク制御ユニット１００のＤＫＣ内結合部３を相互に結合するＳＷ（スイッチ）５（第一の結合部）を備える。各ディスク制御装置１００は、ＳＷ５を介し、互いのキャッシュメモリ４にアクセスし、データをリード／ライトするなどの通常のアクセス処理を実行する。

【００１９】なお、ＳＷ５は、ＬＳＩ等で構成されるメモリパス用スイッチで構成してもよい。この場合、ディスク制御装置２００内の各ディスク制御ユニット１００に共通の電源ボックスから給電するとした場合に用いられる。ＬＳＩ等で構成されるメモリパス用スイッチで構成することで安価にできる。

【００２０】あるいは、ＳＷ５は、データ伝送用のケーブルで構成してもよい。この場合、ディスク制御ユニット毎に電源ボックスを持たせることで、ディスク制御装置２００内の各ディスク制御ユニット１００の電源を独立して給電するとした場合に用いられる。各ディスク制御ユニット１００の電源を独立して給電することで、電源断に耐え得る構成とできる。

【００２１】さらに、ディスク制御システム１０００において、複数のディスク制御装置２００に跨り、各ディスク制御ユニット１００のＤＫＣ内結合部３を相互に結合するＳＷ（スイッチ）６（第二の結合部）を備える。各ディスク制御ユニット１００は、ＳＷ６を介し、互いのキャッシュメモリ４にアクセスし、データの転送を実行する。なお、ＳＷ６はメモリパス用スイッチで構成してもよい。この場合、ディスク制御システム１０００全体におけるディスク制御装置２００内の各ディスク制御ユニット１００に共通の電源ボックスから給電するとした場合に用いられる。ＬＳＩ等で構成されるメモリパス用スイッチで構成することで安価にできる。

【００２２】なお、変形例として、各ディスク制御ユニット１００のＤＫＣ内結合部３をＳＷ（スイッチ）６でもって結合するのではなく、各ディスク制御装置２００に跨り、各ディスク制御ユニット１００のＳＷ５を相互に接続し、通常のアクセスやデータ転送の処理を実行できるようにしてもよい。この場合、ＳＷ（スイッチ）６を省略でき、システム構成の簡素化が図れる。

【００２３】ここで、前述した、ＤＫＣ内結合部（内部結合部）３、ＳＷ（第一の結合部、密結合）５、及びＳＷ（第二の結合部、疎結合）６の構成に関し、二つの事例を用いてより具体的に説明する。

【００２４】＝＝＝＝事例１＝＝＝＝

本事例１は、ＳＷ５及びＳＷ６を同一のプロトコルで実現した事例である。なおかつ、ＳＷ５及びＳＷ６はＤＫＣ内結合部３を拡張した構成としている。

【００２５】まず、ＬＳＩで実現したＤＫＣ内結合部３のブロック図を図２に示す。図２に示すように、ＤＫＣ内結合部３は、セレクタ部３ａとパス制御部３ｂ乃至３ｉとを備える。このセレクタ部３ａに対してパス制御部３ｂ乃至３ｉが接続されている。パス制御部３ｂ，３ｃは、図１に示される各電源クラスタＡ，Ｂ双方のチャネル側御部１の接続パスと接続されている。パス制御部３ｄ，３ｅは、図１に示される各電源クラスタＡ，Ｂ双方のディスク制御部２の接続パスと接続されている。パス制御部３ｆは、ＤＫＣ内結合部３の属する電源クラスタＡあるいは電源クラスタＢのＣＭ４の接続パスと接続されている。パス制御部３ｇ，３ｈは、図１に示される各電源クラスタＡ，Ｂ側双方のＳＷ５の接続パスと接続されている。パス制御部３ｉは、図１に示されるＤＫＣ内結合部３の属する電源クラスタＡあるいは電源クラスタＢの側のＳＷ６の接続パスと接続されている。このＤＫＣ内結合部３の動作については後述する。

【００２６】次に、ＬＳＩで構成したＳＷ５のブロック図を図３に示す。なお、ＳＷ６のハードウエア構成も図３のＳＷ５と同様である。図１に示すように、ＳＷ５は、二つのディスク制御ユニット１００における各電源クラスタＡ，ＢのＤＫＣ内結合部３、即ち、計４つのＤＫＣ内結合部３と接続している。したがって、図３のブロック図では、４ポートの入出力を有するＳＷ５の事例が示される。ＳＷ５は、４つの制御部５ａ乃至５ｄと、受信部５ｅ乃至５ｈと、送信部５ｉ乃至５ｌとを備え

る。各受信部５ｅ乃至５ｈと各制御部５ａ乃至５ｄとは、リクエスト線及びグラントＩＤ線Req/Gntを含んだデータ線で相互に接続されている。また、各制御部５ａ乃至５ｄは、それぞれ対応する各送信部５ｉ乃至５ｌと接続されている。各受信部５ｃ乃至５ｈ及び各送信部５ｉ乃至５ｌは、それぞれバッファを備え、相互に接続されている。

【００２７】次に、ケーブルで構成したＳＷ５のブロック図を図４に示す。図４では、図１においてＳＷ５で示された部分をケーブルとして結線した構成を示している。

【００２８】以上、説明したＳＷ５の動作について、図５のフローチャートを参照して説明する。なお、本明細書のフロチャートにおいて”Ｓ”はステップ（工程）を意味する。図５には、ＳＷ５を介し、図１に示される二つのディスク制御ユニット１００（ＤＫＣ＃０及びＤＫＣ＃１）のＤＫＣ内結合部３間においてデータ及びコマンド等が送受される様子が示される。概念としては、アクセス先のアドレス設定などをディスク制御ユニット１００側で行い、直接にアクセス先のアドレスを指定してアクセスする。

【００２９】具体的には、ＤＫＣ＃０が送信したReadコマンドをＳＷ５はＤＫＣ＃１へ送信する（Ｓ１００）。このReadコマンドのデータは、図６（ａ）に示すように、転送先ＣＭアドレス、転送元ＣＭアドレス、転送長及びコマンドとしてのReadで構成される。次いで、ＤＫＣ＃１のＣＭ４にアクセスがあると、ＤＫＣ＃１はデータとステータスを順に送信する。これらデータとステータスをＳＷ５はＤＫＣ＃０へ送信する（Ｓ１１０、Ｓ１２０）。このデータには、図６（ｂ）に示すように、転送先ＣＭアドレス、転送元ＣＭアドレス及び転送長が付帯する。

【００３０】一方、ＤＫＣ＃０がWriteコマンドとデータを順に送信すると、ＳＷ５は、これらWriteコマンドとデータをＤＫＣ＃１へ送信する（Ｓ１３０、Ｓ１４０）。これらWriteコマンド及びデータのデータ構造は図６（ａ）（ｂ）に示すものと同様であり、図６（ａ）におけるコマンドとしてのReadがWriteとなる。これWriteコマンドとデータを受信したＤＫＣ＃１のＣＭ４にアクセスがあると、ＤＫＣ＃１はステータスを送信する。このステータスをＳＷ５はＤＫＣ＃０へ送信する（Ｓ１５０）。このステータスのデータ構造は図６（ｃ）に示すものと同様である。

【００３１】次に、前述したＳＷ６の動作について、図７のフローチャートを参照して説明する。図７には、ＳＷ６を介し、図１に示される二つのディスク制御装置２００内のディスク制御ユニット１００（ＤＫＣ＃０及びＤＫＣ＃２）のＤＫＣ内結合部３間においてデータ及びコマンド等が転送される様子が示される。概念としては、図１のチャネル制御部１、ディスク制御部２及びＣＭ４のそれぞれには、その機能を実現するためのプロセッサを備えている。そして、ＣＭ４のアドレス管理を行うべく、データ転送に先立ち、ＤＫＣ＃０及びＤＫＣ＃２のプロセッサ間の通信において、アクセス先のＤＫＣに対してアドレスの設定等を要求し、アクセス先のアドレスを取得する。そして、取得したアクセス先のアドレスを指定してデータ転送を実行する。

【００３２】具体的には、ＤＫＣ＃０がデータ転送の要求コマンドを発行し、この要求を受けたＳＷ６はＤＫＣ＃２へ転送する（Ｓ２００）。この要求コマンドは、図８（ａ）に示すように、転送先プロセサを指定するアドレス、転送元プロセサを指定するアドレス、転送長、及びコマンドとしての転送要求で構成される。次いで、ＤＫＣ＃２のＣＭ４のアクセスに必要なアドレスを算出し、算出したアドレスを転送許可と共に送信する。これらアドレス及び転送許可を受けたＳＷ６はＤＫＣ＃０へ転送する（Ｓ２１０）。この転送許可のコマンドは、図８（ｂ）に示すように、転送先プロセサアドレス、転送元プロセサアドレス、転送長、及びコマンドとしての転送許可で構成される。次いで、アドレス及び転送許可を受信したＤＫＣ＃０は、Writeコマンドとデータを順に送信する。すると、ＳＷ６は、これらWriteコマンドとデータをＤＫＣ＃２へ送信する（Ｓ２２０、Ｓ２３０）。これらWriteコマンド及びデータのデータ構造は図６（ａ）（ｂ）に示すものと同様であり、図６（ａ）におけるコマンドとしてのReadがWriteとなる。これWriteコマンドとデータを受信したＤＫＣ＃２のＣＭ４にアクセスがあると、ＤＫＣ＃２はステータスを送信する。このステータスをＳＷ６はＤＫＣ＃０へ転送する（Ｓ２４０）。このステータスのデータ構造は図６（ｃ）に示すものと同様である。

【００３３】＝＝＝＝事例２＝＝＝＝

本事例２では、ＳＷ６については、事例１とは異なり、ＳＷ５と異なる別のプロトコルでもって構成する。すなわち、ＳＷ６は、例えばホストチャネル同様の接続とし、ファイバチャネル上でマッピングしたＳＣＳＩコマンド等により論理アドレスでアクセスすることで実現する。一方、ＳＷ５については、事例１同様に、ＤＫＣ内結合部３を拡張した構成であり、動作も事例１と同様である。したがって、事例１と相違するＳＷ６の構成及び動作を中心に説明する。

【００３４】具体的な構成としては、図９のブロック図に示すように、事例１の場合を示す図２のブロック図に比し、パス制御部３ｉとＳＷ６接続パスとの間にプロトコル変換部６が挿入されている。この点以外は、前述した図２の場合と同様であるため、図２と相違するプロトコル変換部６について説明する。プロトコル変換部６は、図９に示すように、プロセッサ７ａ、メモリ７ｂ、パス制御部７ｃ，７ｄ、バッファ７ｅ，７ｆ、及びパケット変換部７ｇ，７ｈで構成される。プロセッサ７ａ、

メモリ７ｂ、パス制御部７ｃ，７ｄ、及びパケット変換部７ｇ，７ｈは共通のバスに接続される。図９に示すように、メモリ７ｂを適宜使用するプロセッサ７ａの制御の下、パス制御部７ｃ、バッファ７ｅ、パケット変換部７ｇ、パス制御部７ｄといった順序でプロトコルが変換され、データがＤＫＣ内結合部３からＳＷ６へ送信される。反対に、メモリ７ｂを適宜使用するプロセッサ７ａの制御の下、パス制御部７ｄ、バッファ７ｆ、パケット変換部７ｈ、パス制御部７ｃといった順序でプロトコルが変換され、データがＳＷ６経由でＤＫＣ内結合部３へ転送される。

【００３５】次に、前述したＳＷ６の動作について、図１０のフローチャートを参照して説明する。図１０には、ＳＷ６を介し、図１に示される二つのディスク制御装置２００内のディスク制御ユニット１００（ＤＫＣ＃０及びＤＫＣ＃２）のＤＫＣ内結合部３間においてデータ及びコマンド等が転送される様子が示される。ＤＫＣ＃０は、Writeコマンドとデータを順に送信する。すると、ＳＷ６は、これらWriteコマンドとデータをＤＫＣ＃２へ送信する（Ｓ３００、Ｓ３１０）。このデータアクセスを受けたＤＫＣ＃２は、ＳＷ６経由でステータスを送信する（Ｓ３２０）。

【００３６】これらWriteコマンド、データ及びステータスのデータ構造を図１１（ａ）（ｂ）（ｃ）に示す。Writeコマンドは、図１１（ａ）に示すように、転送先ポートアドレス、転送元ポートアドレス、転送長、コマンドとしてのwrite、論理アドレス、及び転送サイズで構成される。データでは、図１１（ｂ）に示すように、転送先ポートアドレス、転送元ポートアドレス、及び転送長が付帯する。ステータスは、図１１（ｃ）に示すように、転送先ポートアドレス、転送元ポートアドレス、転送長、及びステータス情報で構成される。

【００３７】ここで、以上説明した実施例で用いられるＳＷ５の一般的な特性について説明する。図１２（ａ）に示すように、ＳＷ５のポート０乃至３の入出力は一対一の関係にあり、データ転送の効率がよい。一方、図１２（ｂ）に示すように、ＳＷ５のポート０乃至３の入出力が任意の関係の場合では、入力ポート０，１が出力ポート０に対応し、入力ポート２，３が出力ポート２に対応する。この場合、平均的にデータ転送の５０％が論理的に競合するため、ハードウエアの性能を５０％しか活用できない。このため、データ転送の効率が低下した状態となる。すなわち、図１３（ｂ）のグラフに示すように、ＳＷ５に接続されるクラスタ（図１中のディスク制御ユニット１００に相当）の数が増えるほど、効率が低下することとなる。例えば、図１３（ｂ）に示すように複数組のクラスタをＳＷ５に接続すると、効率は５０％となる。つまり、効率が低下すると広帯域化が必要となり、高価格化という問題を招く。

【００３８】そこで、本発明では、図１４のブロック図に示すように、一つのディスク制御装置２００を構成するクラスタ（ディスク制御ユニット１００）の数を二つとし、これら２クラスタ間をＳＷ５で接続するとすれば、図１３（ｂ）のグラフに示すように、クラスタ数を一組（図中のクラスタ数が”１”の場合）とでき、効率を１００％とすることができる。

【００３９】次に、以上説明した、ディスク制御システム１０００、ホストコンピュータ３００、及びディスク装置４００を備えたディスクシステムの全体的な動作について、図１５及び図１６のフロチャートを参照して説明する。適宜、図１のブロック図を参照されたい。なお、図面において”Ｓ”はステップ（工程）を意味する。図１５に示すように、先ず、ホストコンピュータ３００が処理の要求を開始する（Ｓ１０）。ホストコンピュータ３００に接続されたディスク制御ユニット１００は、自己のキャッシュメモリ４のデータに対するアクセスか否かを判別する（Ｓ２０）。この判別の結果、自己のキャッシュメモリ４のデータに対するアクセスであれば、そのアクセスパスが正常か否かを確認する（Ｓ３０）。この確認の結果、アクセスパスが正常であれば、自己のキャッシュメモリ４にアクセスし、データのリード／ライトの処理を実行して終了する（Ｓ４０→Ｓ５０）。

【００４０】一方、Ｓ２０において、自己のキャッシュメモリ４のデータに対するアクセスでない場合（Ｓ２０：ＮＯ）、ディスク制御ユニット１００は、同じディスク制御装置２００内の他方（ペア、電源クラスタＡ，Ｂのうちの他方）のディスク制御ユニット１００に対するアクセスか否かを判断する（Ｓ６０）。この判断の結果、他方のディスク制御ユニット１００に対するアクセスである場合（Ｓ６０：ＹＥＳ）には、ＳＷ５を介し、他方のディスク制御ユニット１００のＤＫＣ内結合部３を部を通じて他方のディスク制御ユニット１００のキャッシュメモリ４にアクセスする（Ｓ７０）。

【００４１】また、Ｓ３０において、アクセスパスが正常でない場合にも、Ｓ７０の処理を実行する。この場合、例えば、電源クラスタＡ側において、チャネル制御部１からＤＫＣ内結合部３を通じたキャッシュメモリ４への通信路に障害が発生した場合、電源クラスタＡのチャネル制御部１が電源クラスタＢのＤＫＣ内結合部３へ接続する。そして、このＤＫＣ内結合部３が電源クラスタＡ側のＳＷ５を介して電源クラスタＡ内のＤＫＣ内結合部３経由で、電源クラスタＡ内のキャッシュメモリ４にアクセスする。このような迂回ルートを有することで対障害性を向上できる。

【００４２】一方、Ｓ６０における判断の結果、他方のディスク制御ユニット１００に対するアクセスでない場合（Ｓ６０：ＮＯ）には、図１６のＢの処理に移り、他のディスク制御装置２００へのデータアクセスと判断し（Ｓ８０）、ＳＷ６を介し、ＤＫＣ内結合部３経由でキ

ャッシュメモリ４へデータアクセスし、データの転送を行う。

【００４３】また、本実施の形態にあっては、各ディスク制御装置の内部において、各ディスク制御ユニットの内部結合部を相互に結合する第一の結合部により、データをリード／ライトする。なおかつ、複数のディスク制御装置に跨り、各ディスク制御ユニットの内部結合部を相互に結合する第二の結合部によって、データを転送する。

【００４４】第一の結合部による密な結合でもってデータをリード／ライトするとともに、第二の結合部による疎な結合でもってデータを転送する。このような、役割分担された結合方式により、高価格化を招くことなくスケーラビリティを向上できる。

【００４５】

【発明の効果】低価格化を維持しながらも、データの転送やリード／ライトの処理の効率並びにスケーラビリティを向上できる。

【図面の簡単な説明】

【図１】　本発明の一実施の形態であるディスクシステムの構成を示すブロック図である。

【図２】　本発明の一実施の形態に係るＤＫＣ内結合部３の一構成例を示すブロック図である。

【図３】　本発明の一実施の形態に係るＳＷ５の一構成例を示すブロック図である。

【図４】　本発明の一実施の形態に係るＳＷ５の他の構成例を示すブロック図である。

【図５】　本発明の一実施の形態に係る二つのディスク制御ユニット１００間のデータ及びコマンド等がＳＷ５を介して送受される様子を示すフローチャートである。

【図６】　本発明の一実施の形態に係るＳＷ５が受送信するデータ構造の例を示す図表である。

【図７】　本発明の一実施の形態に係る二つのディスク装置２００間のデータ及びコマンド等がＳＷ６を介して送受される様子を示すフローチャートである。

【図８】　本発明の一実施の形態に係るＳＷ６が転送するデータ構造の例を示す図表である。

【図９】　本発明の一実施の形態に係るＤＫＣ内結合部３及びプロトコル変換部７の一構成例を示すブロック図である。

【図１０】　本発明の一実施の形態に係る二つのディスク装置２００間のデータ及びコマンド等がＳＷ６を介して送受される様子を示すフローチャートである。

【図１１】　本発明の一実施の形態に係るＳＷ６が転送するデータ構造の例を示す図表である。

【図１２】　本発明の一実施の形態及び従来の技術に用いられるＳＷ５の一般的な特性を示す模式図である。

【図１３】　本発明の一実施の形態及び従来の技術に用いられるＳＷ５を示し、（ａ）はその接続構成を示すブロック図であり、（ｂ）は接続されるクラスタ数に応じた効率を示すグラフである。

【図１４】　本発明の一実施の形態に係るディスク制御ユニット（クラスタ）の接続形態を示すブロック図である。

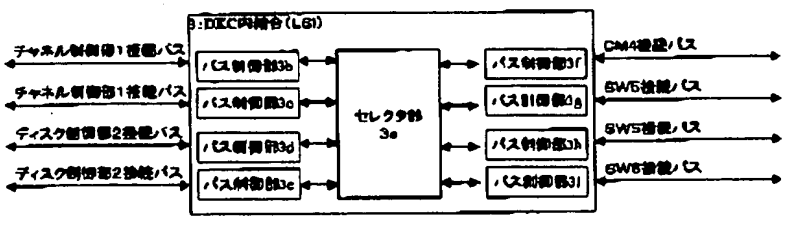
【図１５】　本発明の一実施の形態に係るディスクシステムの全体的な動作を示すフローチャートである。

【図１６】　本発明の一実施の形態に係るディスクシステムの動作の一部を示すフローチャートである。

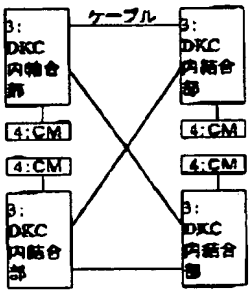
【符号の説明】

１	チャネル制御部
２	ディスク制御部
３	ＤＫＣ内結合部
４	キャッシュメモリ（ＣＭ）
５	ＳＷ（第一の結合部、密結合）
６	ＳＷ（第二の結合部、疎結合）
７	プロトコル変換部
１００	ディスク制御ユニット
２００	ディスク制御装置
３００	ホストコンピュータ
４００	ディスク装置
１０００	ディスク制御システム

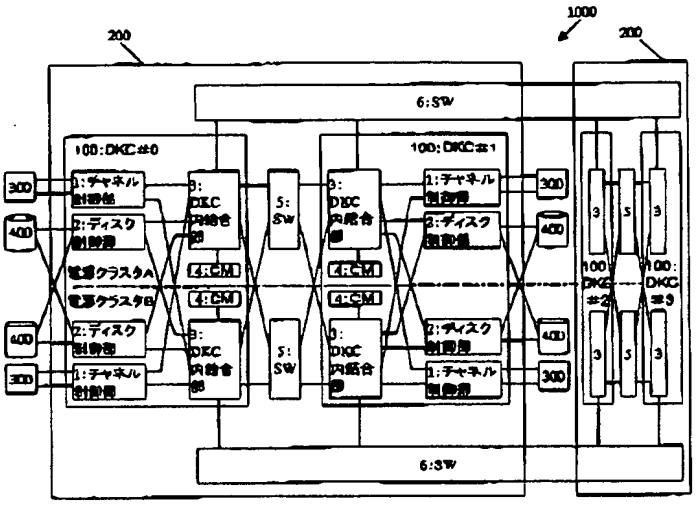
【図２】



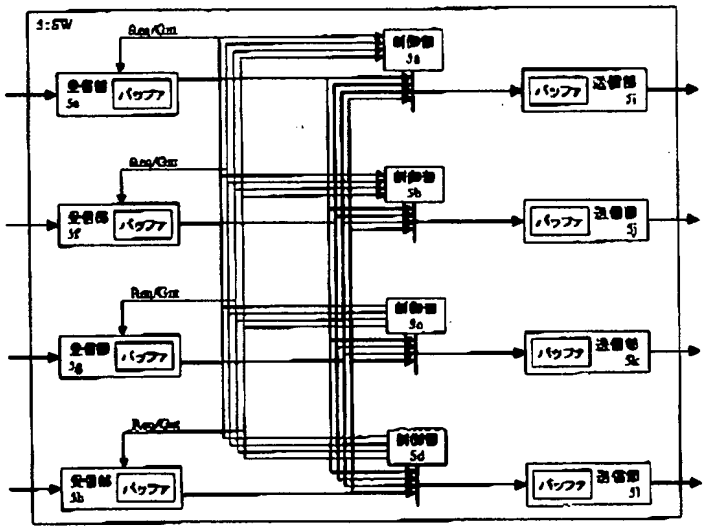
【図４】



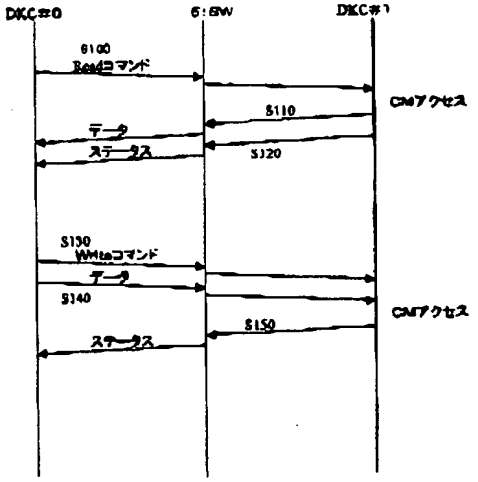
【図１】



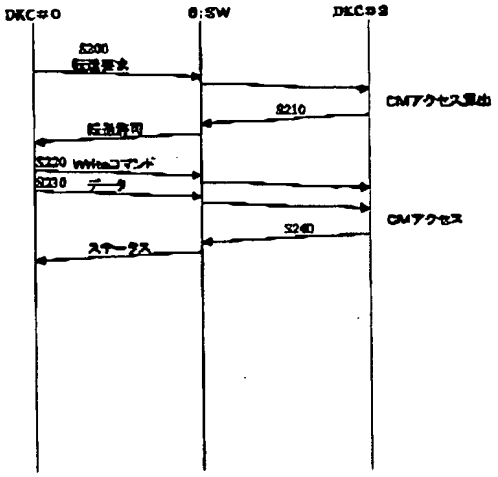
【図３】



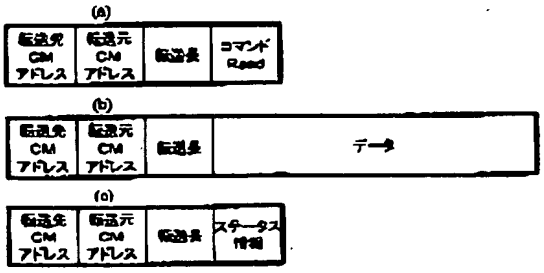
【図５】



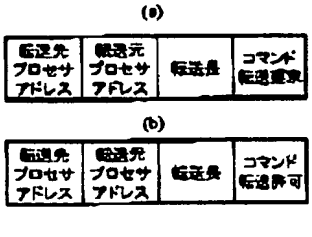
【図７】



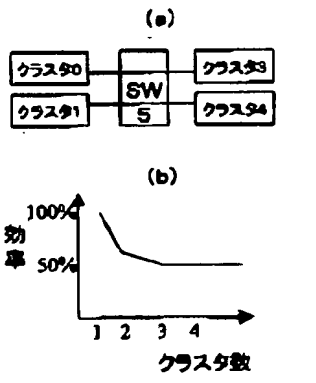
【図６】



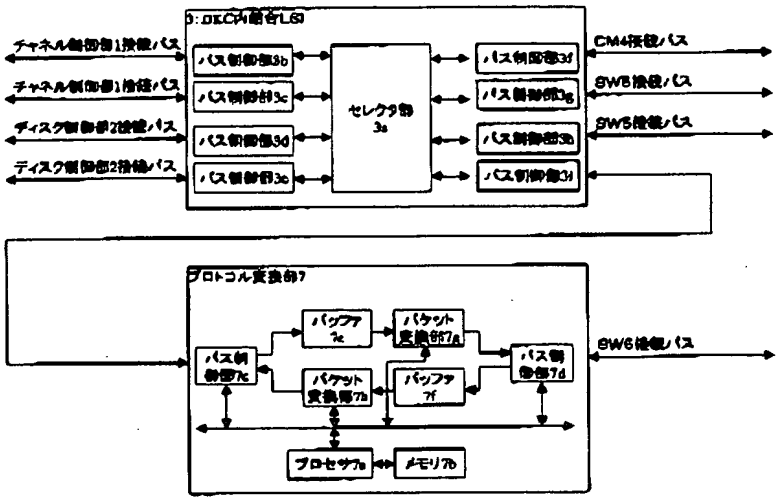
【図８】



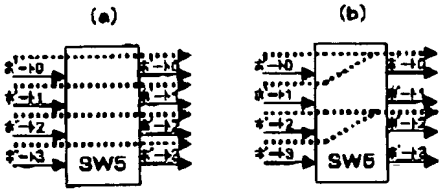
【図１３】



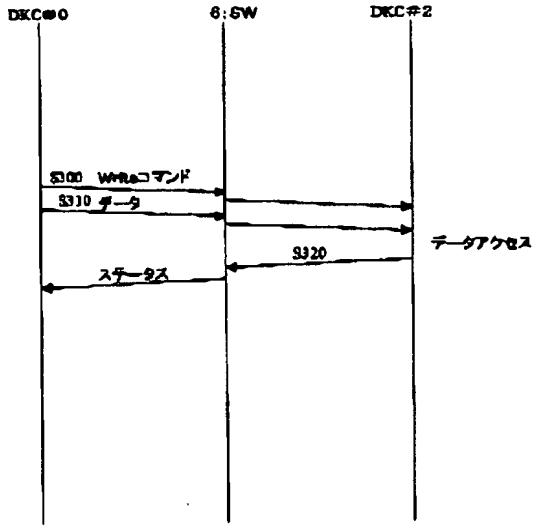
【図９】



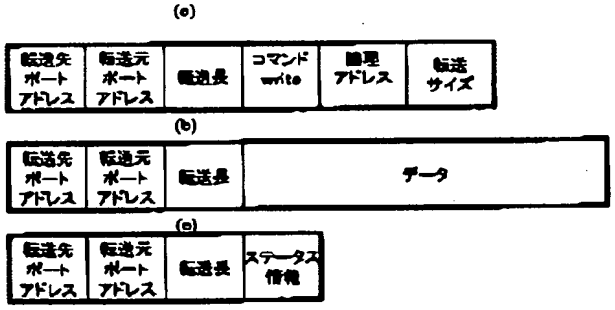
【図１２】



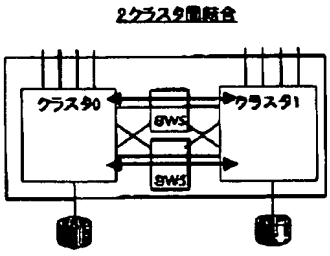
【図１０】



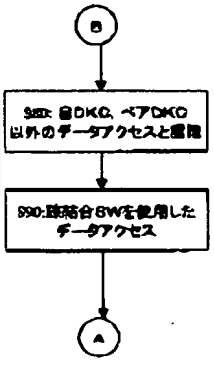
【図１１】



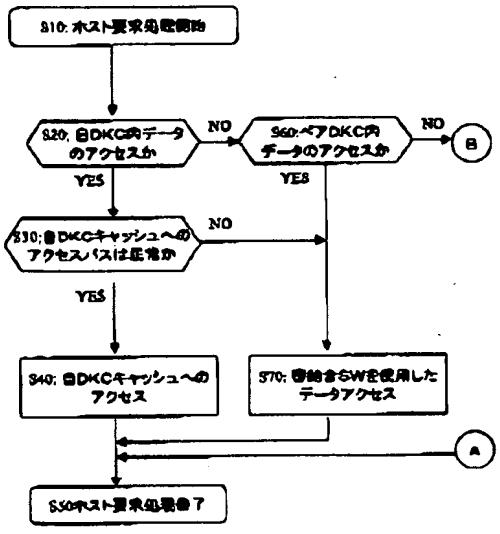
【図１４】



【図１６】



【図１５】



フロントページの続き

(51)Int.Cl.'	識別記号	FI	ﾃｰﾏｺｰﾄﾞ(参考)
G06F 13/12	310	G06F 13/12	310E

Fターム(参考) 5B005 JJ12 MM11 NN75
5B014 EB05 GA13 GA25 GA26 GA47
5B065 BA01 CA07 CE12 CH01 CH11
ZA13

【公報種別】特許法第１７条の２の規定による補正の掲載
【部門区分】第６部門第３区分
【発行日】平成17年9月22日(2005.9.22)

【公開番号】特開2003-323261(P2003-323261A)
【公開日】平成15年11月14日(2003.11.14)
【出願番号】特願2002-126885(P2002-126885)
【国際特許分類第７版】
ＧＯ６Ｆ 3/06
ＧＯ６Ｆ 12/08
ＧＯ６Ｆ 13/12
【ＦＩ】
ＧＯ６Ｆ 3/06 ３０１Ｂ
ＧＯ６Ｆ 3/06 ３０２Ａ
ＧＯ６Ｆ 3/06 ３０２Ｂ
ＧＯ６Ｆ 12/08 ５０１Ｅ
ＧＯ６Ｆ 12/08 ５５７
ＧＯ６Ｆ 13/12 ３１０Ｅ

【手続補正書】
【提出日】平成17年4月7日(2005.4.7)
【手続補正１】
【補正対象書類名】明細書
【補正対象項目名】特許請求の範囲
【補正方法】変更
【補正の内容】
【特許請求の範囲】
【請求項１】
複数のディスク制御ユニットを有するディスク制御装置を複数備えたディスク制御システムにおいて、
前記ディスク制御ユニットは、
ホストコンピュータとのインターフェースを有する一または複数のチャネル制御部と、
ディスク装置とのインターフェースを有する一または複数のディスク制御部と、
前記ディスク装置にリード／ライトされるデータを一時的に格納するキャッシュメモリ部と前記チャネル制御部と前記ディスク制御部とを相互に接続する内部結合部と、
を備えており、
前記各ディスク制御装置の内部において、データをリード／ライトすべく、前記各ディスク制御ユニットの前記内部結合部を相互に結合する第一の結合部と、
複数の前記ディスク制御装置に跨り、データを転送すべく、前記各ディスク制御ユニットの前記内部結合部を相互に結合する第二の結合部と、
を備えたことを特徴とするディスク制御システム。
【請求項２】
前記ディスク制御装置は二つの前記ディスク制御ユニットを有しており、前記第一の結合部は、該二つのディスク制御ユニットの前記内部結合部を相互に結合することを特徴とする請求項１に記載のディスク制御システム。
【請求項３】
前記第一の結合部又は前記第二の結合部は、メモリパス用スイッチで構成されることを特徴とする請求項１に記載のディスク制御システム。
【請求項４】
前記第一の結合部は、データ伝送用のケーブルで構成されることを特徴とする請求項１

に記載のディスク制御システム。
【請求項５】
前記各ディスク制御装置の内部において、共通の電源から給電される前記各ディスク制御ユニットを前記第一の結合部は結合することを特徴とする請求項１に記載のディスク制御システム。
【請求項６】
複数の前記ディスク制御装置に跨り、データを転送すべく、前記各ディスク制御ユニットの前記第一の結合部を相互に接続することを特徴とする請求項１に記載のディスク制御システム。
【請求項７】
請求項１に記載のディスク制御システムと、前記ディスク制御システムの前記チャネル制御部に対して接続される、データの授受用のホストコンピュータとを備えることを特徴とするディスクシステム。
【請求項８】
請求項１に記載のディスク制御システムと、前記ディスク制御システムの前記ディスク制御部に対して接続される、データ格納用のディスク装置とを備えることを特徴とするディスクシステム。
【請求項９】
複数のディスク制御ユニットを有するディスク制御装置において、
前記ディスク制御ユニットは、
ホストコンピュータとのインターフェースを有する一または複数のチャネル制御部と、
ディスク装置とのインターフェースを有する一または複数のディスク制御部と、
前記ディスク装置にリード／ライトされるデータを一時的に格納するキャッシュメモリ部と前記チャネル制御部と前記ディスク制御部とを相互に接続する内部結合部と、
を備えており、
前記各ディスク制御装置の内部において、データをリード／ライトすべく、前記各ディスク制御ユニットの前記内部結合部を相互に結合する第一の結合部を備えて、
前記各ディスク制御ユニットの前記内部結合部は、複数の前記ディスク制御装置に跨り、データを転送すべく、第二の結合部によって、相互に結合されることを特徴とするディスク制御装置。
【請求項１０】
前記ディスク制御装置は二つの前記ディスク制御ユニットを有しており、前記第一の結合部は、該二つのディスク制御ユニットの前記内部結合部を相互に結合することを特徴とする請求項９に記載のディスク制御装置。
【請求項１１】
前記第一の結合部又は前記第二の結合部は、メモリパス用スイッチで構成されることを特徴とする請求項９に記載のディスク制御装置。
【請求項１２】
前記第一の結合部は、データ伝送用のケーブルで構成されることを特徴とする請求項９に記載のディスク制御装置。
【請求項１３】
共通の電源から給電される前記各ディスク制御ユニットを前記第一の結合部は結合することを特徴とする請求項９に記載のディスク制御装置。
【請求項１４】
複数の前記ディスク制御装置に跨り、データを転送すべく、前記各ディスク制御ユニットの前記第一の結合部が相互に接続されることを特徴とする請求項９に記載のディスク制御装置。
【請求項１５】
ホストコンピュータとのインターフェースを有する一または複数のチャネル制御部と、ディスク装置とのインターフェースを有する一または複数のディスク制御部と、前記ディ

スク装置にリード／ライトされるデータを一時的に格納するキャッシュメモリ部と前記チャネル制御部と前記ディスク制御部とを相互に接続する内部結合部とを備えたディスク制御ユニットを複数有するディスク制御装置におけるデータ通信の制御方法であって、

前記各ディスク制御装置の内部において、前記各ディスク制御ユニットの前記内部結合部を相互に結合する第一の結合部により、データをリード／ライトするとともに、

複数の前記ディスク制御装置に跨り、前記各ディスク制御ユニットの前記内部結合部を相互に結合する第二の結合部によって、データを転送することを特徴とするディスク制御装置におけるデータ通信の制御方法。

【請求項１６】

前記ディスク制御装置は二つの前記ディスク制御ユニットを有しており、前記第一の結合部は、該二つのディスク制御ユニットの前記内部結合部を相互に結合することを特徴とする請求項１５に記載のディスク制御装置におけるデータ通信の制御方法。

【請求項１７】

複数のディスク制御ユニット及び少なくとも一つの第一の結合部を含む複数のディスク制御装置と、前記ディスク制御装置間に設けられる少なくとも一つの第二の結合部と、を備えるディスク制御システムであって、

前記各ディスク制御ユニットは、ホストコンピュータとのインターフェースを有する少なくとも一つのチャネル制御部と、ディスク装置とのインターフェースを有する少なくとも一つのディスク制御部と、前記チャネル制御部と前記ディスク制御部と前記ディスク装置にリード／ライトされるデータを一時的に格納するキャッシュメモリとを相互に接続する内部結合部と、を含み、

前記各ディスク制御装置の前記第一の結合部は、前記各ディスク制御装置の内部においてデータをリード／ライトすべく、前記各ディスク制御装置を構成する各ディスク制御ユニットの前記内部結合部を相互に接続し、

前記第二の結合部は、前記各ディスク制御装置を跨りデータを転送すべく、前記すべてのディスク制御装置における前記ディスク制御ユニットの前記内部結合部を相互に接続する

ことを特徴とするディスク制御システム。

【請求項１８】

ディスク制御システムと、データを授受するために前記ディスク制御システムと接続される少なくとも一つのホストコンピュータと、を備えるディスクシステムであって、

前記ディスク制御システムは、複数のディスク制御ユニット及び少なくとも一つの第一の結合部を含む複数のディスク制御装置と、前記各ディスク制御装置を跨りデータ転送を行う少なくとも一つの第二の結合部と、を備え、

前記各ディスク制御ユニットは、ホストコンピュータと通信可能なインターフェースを有する少なくとも一つのチャネル制御部と、ディスク装置と通信可能なインターフェースを有する少なくとも一つのディスク制御部と、前記チャネル制御部と前記ディスク制御部と前記ディスク装置にリード／ライトされるデータを一時的に格納するキャッシュメモリとを相互に接続する内部結合部と、を含み、

前記各ディスク制御装置の前記第一の結合部は、前記各ディスク制御装置の内部においてデータをリード／ライトすべく、前記各ディスク制御装置のディスク制御ユニットの前記内部結合部の間でデータの転送を行い、

前記第二の結合部は、前記各ディスク制御装置に跨りデータを転送すべく、前記各ディスク制御装置に設けられる全ての前記ディスク制御ユニットの前記内部結合部の間でデータの転送を行い、

前記少なくとも一つのホストコンピュータは、前記いずれかのディスク制御ユニットのチャネル制御部を介して前記ディスク制御システムに接続される

ことを特徴とするディスクシステム。

【請求項１９】

ディスク制御システムと、データを格納するために前記ディスク制御システムと接続さ

れる少なくとも一つのディスク装置と、を備えるディスクシステムであって、
前記ディスク制御システムは、複数のディスク制御ユニット及び少なくとも一つの第一の結合部を含む複数のディスク制御装置と、前記ディスク制御装置間に設けられる少なくとも一つの第二の結合部と、を備え、
前記各ディスク制御ユニットは、ホストコンピュータとのインターフェースを有する少なくとも一つのチャネル制御部と、前記ディスク装置とのインターフェースを有する少なくとも一つのディスク制御部と、前記チャネル制御部と前記ディスク制御部と前記ディスク装置にリード／ライトされるデータを一時的に格納するキャッシュメモリとの間でデータ接続を行う内部結合部と、を含み、
前記第一の結合部は、同一のディスク制御装置内に設けられる前記各ディスク制御ユニットの前記内部結合部を相互に接続して、そのディスク制御装置の内部においてデータをリード／ライトし、
前記第二の結合部は、前記各ディスク制御装置にそれぞれ設けられる前記ディスク制御ユニットの前記内部結合部を相互に接続して、前記ディスク制御装置に跨りデータを転送し、
前記少なくとも一つのディスク装置は、前記いずれかのディスク制御ユニットのディスク制御部を介して前記ディスク制御システムに接続される
ことを特徴とするディスクシステム。
【請求項２０】
複数のディスク制御ユニットと、少なくとも一つの第一の結合部と、を備えるディスク制御装置であって、
前記各ディスク制御ユニットは、ホストコンピュータとのインターフェースを有する少なくとも一つのチャネル制御部と、ディスク装置とのインターフェースを有する少なくとも一つのディスク制御部と、前記チャネル制御部と前記ディスク制御部と前記ディスク装置にリード／ライトされるデータを一時的に格納するキャッシュメモリとを相互に接続する内部結合部と、を含み、
前記第一の結合部は、前記ディスク制御装置の内部においてデータをリード／ライトすべく、前記ディスク制御装置に設けられる前記各ディスク制御ユニットの前記内部結合部間におけるデータ接続を行い、
前記各ディスク制御ユニットの前記内部結合部は、複数のディスク制御装置に跨りデータを転送すべく、少なくとも一つの第二の結合部を介して、他のディスク制御装置の少なくとも一つの内部結合部に接続される
ことを特徴とするディスク制御装置。
【請求項２１】
ホストコンピュータとのインターフェースを有する少なくとも一つのチャネル制御部と、ディスク装置とのインターフェースを有する少なくとも一つのディスク制御部と、前記チャネル制御部と前記ディスク制御部と前記ディスク装置にリード／ライトされるデータを一時的に格納するキャッシュメモリとを相互に接続する内部結合部と、を含む複数のディスク制御ユニットを備えるディスク制御装置におけるデータ通信の制御方法であって、
前記各ディスク制御ユニットの前記内部結合部を相互に接続する第一の結合部を用いて前記ディスク制御装置の内部においてデータをリード／ライトし、
前記各ディスク制御ユニットの前記内部結合部を、他のディスク制御装置の少なくとも一つの内部結合部に接続する第二の結合部を用いて、複数のディスク制御装置を跨りデータを転送する
ことを特徴とするデータ通信の制御方法。